

A BENCHMARKING EXERCISE ON LATIN AMERICAN WATER UTIL

Carlos A. Romero and Gustavo Ferro

Working Paper N° 27

ISBN: 978-987-519-139-6

(August 2008)

CEER

Centro de Estudios Económicos de la Regulación

Universidad Argentina de la Empresa

Lima 717

C1053AAO Buenos Aires, Argentina

Telephone / Fax: 54-11-4000-7693

E-mail: ceer@uade.edu.ar

<http://www.uade.edu.ar>

Romero, Carlos

A Bench marking exercise on latin American water utilities / Carlos Romero y Gustavo Ferro. - 1a ed. - Buenos Aires : Ediciones UADE - Universidad Argentina de la Empresa, 2008.

23 p. ; 27x21 cm. - (Working Paper; 27)

ISBN 978-987-519-139-6

1. Economía. I. Ferro, Gustavo II. Título
CDD 330

Fecha de catalogación: 19/12/2008

A BENCHMARKING EXERCISE ON LATIN AMERICAN WATER UTILITIES

Carlos A. Romero¹ and Gustavo Ferro²

Resumen

El objetivo de este estudio es estimar funciones de frontera de costo tanto estocásticas como de programación matemática, partiendo de la base de datos de ADERASA. Esta es una asociación de reguladores de agua y saneamiento de América Latina que ha hecho un esfuerzo sistemático de recolección de datos. Técnicas econométricas y de DEA fueron empleadas aquí. Este estudio llena una brecha en la comprensión de la eficiencia relativa en la prestación del servicio de agua y saneamiento en América Latina. Primero, presentamos un recorrido por la literatura empírica relativa a las fronteras de costo y producción en el sector de agua y saneamiento. Dicho recorrido facilita la determinación de las variables a escoger en los modelos a ser estimados, al arrojar luz sobre el ambiente de operación. Segundo, algunos modelos son estimados, siendo diferentes unos de otros en la especificación y en las variables ambientales incluidas. Los resultados son satisfactorios, con los signos esperados y valores plausibles para los coeficientes. El trabajo precedente fue completado con un análisis de consistencia entre metodologías y entre resultados al interior de cada método.

Abstract

The aim of this study is to estimate both stochastic and mathematical programming cost functions and frontier studies with the ADERASA's database. ADERASA is a Latin American association of water regulators, which has done a systematic effort on data collection. Econometric and DEA techniques were employed here. This study fills a gap on the understanding of relative efficiency in the Latin American water sector., through a consistent database. Firstly we present a survey of the empirical literature related with cost and production frontiers in the water and sanitation sector. The survey shed light in order to determine the variables to choose in the models to be estimated. Secondly, some models were estimated, differing each other on the specification, and the environmental variables included. The results are satisfactory, with the expected signs and plausible values for the coefficients. Some consistency between methodologies was found.

¹ Economic Regulation Research Center (CEER) and Institute of Economics UADE, Buenos Aires (Argentina) cromero@uade.edu.ar.

² Economic Regulation Research Center (CEER) and Institute of Economics UADE, Buenos Aires (Argentina) gferro@uade.edu.ar, and CONICET.

I. Introduction

In recent years, a regional association of Latin American water and sanitation regulators (ADERASA) has done a great effort of data collection in order to develop benchmark studies. The database in use was compiled by ADERASA and it contains information for Latin American water and sewerage firms in the years 2003-2005.

The aim of this study is to estimate both stochastic and mathematical programming cost functions and frontier studies with the data collected by ADERASA. Econometric and DEA techniques were employed to the calculations. The study fills a gap on the understanding of the relative efficiency in Latin American water and sanitation.

The lack of cross-country studies for the region is one of the motivations of the present article. The possibilities open were many, since the recent advances in techniques and methodologies to study relative efficiency. In the past, the greater obstacle to the development of this kind of studies was the absence of a systematic database from many countries and many firms. With the third wave of the results of the ADERASA survey, the number of firms and operators surveyed is sufficient to develop a cross sectional study for the year 2005. A panel data analysis is yet impossible, since the incompleteness of the data from 2003 and 2004; either in observations and in blanks in particular items of the survey. The possibility is open when 2006 data become available.

A detailed overview of the literature was done, to recognize the set of variables that have been included in previous studies. In the empirical task, all the included variables in the previous work were tested, in one way or in other. From the survey it was also decided to do the estimations at Purchasing Power Parity (PPP) prices, in order to avoid artificial differences in costs and prices when estimating cost functions with firms from many countries. The survey also was useful to think other variables not previously studied and to apply them to the estimates. In the survey it was applied a double criteria: studies were grouped by countries or regions, and in a chronological order.

Following this introduction, in the second section it was made an exhaustive survey of the empirical literature related with cost and production frontiers in the water and sanitation sector. The survey shed light in order to determine the variables to choose in the model to be estimated. After that, in the third section, descriptive statistics of the database are presented. In the fourth section, some models were estimated, both econometric and DEA, differing each other on the specification, and the environmental variables included. It is also made a consistency study between methodologies and intra-methodology within the different models. In the fifth section the main results are shown. Finally, section six concludes.

II. A brief survey of the empirical literature

On England and Wales, there are found the pioneer studies of Price (1993), estimating Operative Costs (OPEX), Stewart (1993) who developed estimates of water cost functions for the OFWAT system and Stewart (1994) studying sewerage costs for the OFWAT system of the period 1992-93. Bosworth, Stoneman and Thanassoulis (1996) extended the Stewart (1993 and 1994) studies in some directions. They examined the use of cost and production functions, discussed conceptual issues concerning with the functional form to be chosen, measuring problems and the ideal type of basis information. Botasso and Conti (2003) analyzed the evolution of the operative costs efficiency of the English and Welsh sector, estimating a stochastic cost frontier for the period 1995-2001. Saal et al (2004), estimated an input distance function, qualitatively adjusted, with stochastic frontier techniques for the period 1985-2000. Saal and Reid (2004), examined how the regulations – both economic and environmental- have influenced on the productivity growth of the industry of the water and sanitation sector of England and Wales. Saal and Parker (2005) employed an input distance function qualitatively adjusted and stochastic frontier techniques to estimate rates of growth in the productivity operations of the water and sewerage industry of England and Wales.

Fraquelli and Moiso (2005), analyzed the Italian water sector reform, with special emphasis on the cost efficiency frontier of the industry and on the scale economies at the level of “Optimal Territorial Ambits”, in doing so, they estimated a stochastic cost frontier.

The cost frontier for Asia estimated by Estache and Rossi (1999) comprises 50 firms from a database of 1995, provided by the el Asian Development Bank. Estache and Rossi (2002), go further on their previous work, attempting to establish differences of efficiency between private and public operators.

Mobbs and Glennie (2005) run DEA estimates with the ADERASA 2003 database, relating a weighted average of outputs to a weighted average of inputs. Romero (2005) is the more direct precedent to the present work, being used the ADERASA database for efficiency estimates, through cost functions.

Crampes et al (1997) estimated a cost function for the Brazilian water sector. They employed the same variables than Stewart (1993) did. Moreira and Fonseca (2005), suggested evaluation criteria for the productivity estimates that emerged from DEA and stochastic frontier analysis. Tupper and Resende (2003) quantified relative efficiency of water and sanitation firms at a state level in Brazil during the period 1996-2000. Sabbioni (2005) measured the relative performance of water and sanitation public firms in Brazil by means of econometric techniques He applied a cost function approach as the more appropriate, based on the operative environment of operation, the ability to deal with multiple outputs, the absence of endogeneity problems, the availability of information and the technological specification.

Berg and Lin (2005) evaluated the consistency of the performance rankings of public Peruvian firms. The stochastic frontier and the DEA analysis provided similar rankings. Lin (2005) examined how the introduction of quality variables affects the comparisons between public firms in Peru.

Estache and Kouassi (2002) analyzed the determinants of the levels of efficiency reached by 21 African public operators, using an estimation of production frontier for the sector.

III. Empirical analysis of data

The database in use was compiled by ADERASA and it contains information for Latin American water and sewerage firms in the years 2003-2005. Table 1 displays the number of observations by country and by year. In the years previous to 2005, Brazil's data came from another source (SNIS database). Since 2005, where the number of observations increased notably, the responses from Brazil are also originated in ADERASA's survey.

Table 1: Observations ADERASA's database, by country and year.

Country	2003	2004	2005	Total
Argentina	18	8	8	34
Bolivia	3	1	3	7
Brazil	16	16	5	37
Chile	18	18	18	54
Colombia	0	8	38	46
Costa Rica	1	2	2	5
Ecuador	1	1	1	3
Honduras	1	1	1	3
Mexico	0	0	34	34
Nicaragua	1	1	1	3
Peru	9	9	10	28
Panama	1	1	1	3
Paraguay	1	3	4	8
Uruguay	0	0	1	1
Total	70	69	127	266

Source: Own elaboration

The panel is strongly unbalanced, and the amount of responses from every variable is very heterogeneous. The quality and quantity of information improves notably in 2005, thus descriptive statistics on the database are concentrated in the year 2005, which is the base for later estimations on relative efficiency.

The Table 2 shows the descriptive statistics of the variables to be used in the estimations of efficiency frontiers. Since the main objective of the study is to estimate cost frontiers, salaries and the price of other inputs are critical. Then, the number of valid observations falls from 127 to 70 in 2005. In order to present systematically the discussion, the variables are grouped into the following categories: Costs, Outputs, Inputs, and Environment.

III.1 Costs

In the survey there are two categories of OPEX (jointly from water plus sewerage), both measured in current American Dollars:

- (i) Direct OPEX plus General Expenses, excluding depreciations, interests and indirect taxes (OPEXGE);
- (ii) OPEX: labor costs, fuel, electricity, chemicals, etcetera from water and sewerage (OPEX).

OPEX was used in the estimates. This category represented in 2005 a 62% of OPEXGE in average. OPEX data are more homogeneous.

Table 2: Descriptive statistics of variables in use (ADERASA Database 2005)

Variable	Description	Unit	N	Average	Standard Deviation
L	Staff	Nº	121	575	1005
NETW	Water Network	Km	109	1937	3168
NETS	Sewerage Network	Km	108	1193	1905
CLIW	Water Clients	Nº	127	177847	546139
CLIS	Sewerage Clients	Nº	127	180793	363228
COVW	Population Served with Domiciliary Water Service/Total Population	%	127	90.61	13.13
COVS	Population Served with Sewerage/Total Population	%	127	73.94	27.71
METW	Operative Meters/Water Accounts	%	125	63.83	37.99
COVN	Population Served with Domiciliary Water Service/Water Network	inhabitants/km	109	423	183
POPU	Resident Population	Inhabitants	127	818428	1556509
PROW	Water Production	m3/day	126	218354	487270
TRES	Treated Sewage	m3/day	117	36841	91733
UNDW	Underground Water Sources/Total Water Sources	%	121	38.04	42.80
OPEX	Operational Expenses	us\$ / 1000	121	94866	695819
W	Average Salary	us\$ /year	74	17757	25294
R	Price Index from Other Inputs	us\$ year / 1000 clients	72	81.04	82.53
RESB	Residential Billing/Total Billing	%	104	77.51	13.88
RESW	Residential Water Clients/Total Water Clients	%	90	0.93	9.06
UNAW	Unaccounted for Water/Total Water Production	%	126	38.10	17.26
BREW	Water Main Breaks/Water Network length	Nº/ km	60	2.36	4.19
COMP	Client Complaints	Nº	103	41042	101673

Source: Own elaboration on ADERASA database

III.2 Output

There are three sets of production variables for every service, water and sewerage:

- (i) Those related with clients and connections;
- (ii) Those associated to population served, and
- (iii) Those linked to production level.

Both kinds of clients (CLIW and CLIS) are highly correlated. They are also highly correlated with population coverage measures (COVW and COVS) and with water production (PROW). Since the high correlation between the variables, the estimates were run using CLIW as the output variable.

III.3 Inputs and input prices

Between the main inputs it can be named staff (L), proxy of labor, and mains length of water (NETW) and sewerage (NETS), indicating capital.

The variable L includes all proper staff (normalized to full time equivalent worker) of the operator. There is no detailed information on labor functions. In particular, there are no disposable details on network extension personnel and management or maintenance staff. This is particularly important, because if it is considered only OPEX, those firms that are expanding the networks (where CAPEX are significant) will result relatively penalized in the efficiency measures. They will appear using more L for unit of OPEX than the firms that only operate the service without important CAPEX.

There is no information on proper or outsourced personnel. If a firm outsources an important percentage of its staff, it will suffer a reduction in the efficiency measure calculated. Finally, it is desirable to divide the expenses on personnel onto the two services: water and sanitation, but currently the data is not available.

NETW and NETS are highly correlated, whilst in some observations NETS values are inferior to NETW because of lesser coverage of sewerage.

A cost function relates expenses with outputs and input prices, and additionally it can be included environmental variables. From such cost function it is built the frontier and with this they are derived the individual efficiency measures. Then input prices are critical for the estimations. Since 2005, there is better information on input prices, in particular on wages.

A measure of the average salary was constructed following the practice in the literature, dividing total salaries on staff. That measure was valued at Purchasing Power Parity, as OPEX itself also was.

Because of the lack of more details on input prices, a similar methodology developed by other authors was applied to capital. We named R an index number of the rest of the inputs. The variable was constructed by subtracting from OPEX the labor costs. The result was divided on the number of clients. It also was expressed on purchasing power parity values.

III.4 Environment

The omission of environmental variables distorts the results, whilst to generate them is difficult in a number of situations because of failures on basic information. As environmental variables there were considered:

- (i) Underground Water Volume (UNDW);
- (ii) Unaccounted for Water/Total Water Production (UNAW).

- (iii) Residential Water Clients/Total Water Clients (RESW);
- (iv) Water Main Breaks/Water Network Length (BREW);
- (v) Client complaints (COMP);

All them highlight some kind of environmental condition which could be considerate exogenous. Related with UNDW, those firms with access to underground sources are expected to have lower costs. And the sensitivity of costs with respect to UNAW, RESW, BREWW and COMP are expected to be positive.

The Table 3, shows simple correlation coefficients between environmental variables themselves, and between environmental and OPEX.

Table 3: ADERASA Database, 2005. Simple Correlations Between Environmental Variables

Simple correlation coefficient						
	OPEX	UNDW	COMP	UNAW	RESW	BREW
OPEX	1.000					
UNDW	-0.249	1.000				
COMP	0.493	-0.119	1.000			
UNAW	0.093	0.337	0.116	1.000		
RESW	-0.005	0.236	-0.053	0.063	1.000	
BREW	0.015	-0.183	0.064	0.078	-0.415	1.000

Source: Own elaboration on ADERASA database.

IV. Methodology

IV.1 Model searching: The econometric approach

The open options are to estimate a production function or a cost function (see Sabbioni, 2005, for a comparison of both criteria), and to obtain efficiency scores from econometric or mathematical programming approaches. We choose to estimate a cost function, and we employed both approaches –econometric and mathematical programming-. Moreover, we provide some consistency checks between results of different methodologies.

One advantage of the cost function over the production function approach is the flexibility to adopt different specifications, particularly in the cases when the firm produces more than one product. Moreover, estimation of production function allows obtaining a measure of technical efficiency, but ignores resource allocation problems. Estimation of cost frontiers, on the other hand, gives information on cost differentials due to technical and resource allocation inefficiencies. To separate these two effects it is necessary to formulate some additional assumptions.

The main advantage of non-parametric methods (also known as Data Envelopment Analysis or DEA for short) is that no a priori functional form is imposed to the data. The main disadvantage of this approach is the use of only a subset of the available data for the frontier estimate (those which actually determine the frontier), while the rest of the observations are ignored.

The third methodological decision was to choose a stochastic frontier as the econometric approach since its desirable properties (less sensitivity to outliers, and less arbitrary determination of inefficiency) with respect to the alternative deterministic approach.

In traditional cost analysis the problem faced by the firm is to minimize total costs subject to delivering a given level of output. The solution to this problem generates an optimal set of inputs, which depend on output level and input prices. In the same way, it is possible to estimate the cost function of the firm, which depends only on output level and input prices.

The resulting cost model specification is given by: $C = f(Y, Z, P_L, P_K)$

Where: C: total cost, Y: output, Z: n-dimension vector of other exogenous variables, P_k : price of capital inputs, P_l : price of labor inputs.

The most common specification is the Cobb-Douglas function where the inefficiency terms (ε) enters the model as a multiplicative factor (which turns into additive in the logarithmic form):

$$[1] \quad C = A P_l^{\beta_1} P_k^{\beta_2} Y^{\gamma_0} \prod_i Z_i^{\gamma_i} \exp^{\varepsilon}$$

Applying logarithms to both sides we obtain:

$$[1'] \quad c = \alpha + \beta_1 p_1 + \beta_k p_{2k} + \gamma_0 y + \sum_i \gamma_i z_i + \varepsilon$$

Where β_i and γ_i are parameters to be estimated and small cases represent logarithms of the variables in [1] (levels in capital letters).

The firm with the $\min(\varepsilon_i)$ will be 100% efficient. For this firm ε_i are zero and therefore $\exp(\varepsilon_i)$ equals one. The larger the inefficiency of a particular firm i the term ε_i will be larger and the resulting efficiency measure closer to zero.

In the case of stochastic frontiers, the cost function is similar to the one presented in (1') only that now the error term ε is no longer equal to inefficiency but is decomposed into two terms:

$$[2] \quad \varepsilon_i = u_i + v_i,$$

Where $u_i > 0$ and v_i is not restricted. The v_i term captures the effects of statistical noise and are assumed to be independently and identically distributed with an $N(0, \sigma_v^2)$. The u_i error term represents cost inefficiency and is assumed distributed independently of the v_i and the repressors. Several functional forms have been proposed for the inefficiency term: half-normal, truncated normal, Gamma and exponential. The most common distribution used in empirical tests is the half-normal.

Most water utilities are required to provide services at a fixed tariff, meeting the demand. Since output is exogenous, firms maximize benefits by minimizing costs of producing a given level of output. A cost frontier is then a natural choice, there is also another advantage over the production frontier: it deals better with multiple outputs.

With regard to Z vector, in practice, the costs of regulated public utilities depend on a variety of factors in addition to output levels and input prices. In the comparisons, analysts want the relative ranking to reflect managerial decisions rather than the unique characteristics of service territories beyond managers' control (topography, hydrology and customer density), and historical policies and regulatory decisions.

The effect of exchange rate fluctuations could significantly impact on unit costs and perceived efficiency, masking changes in costs. PPP is one method of correcting this problem. It states that exchange rate between two countries should equal the ratio of the countries' price levels of a fixed basket of goods and services. Purchasing parity prices, are based on a sample basket of goods and services selected to be representative of the GDP for each country. If Purchasing parity prices, specific to water industry expenditure were available they might differ from these, although probably not substantially (OFWAT, 2005).

Assuming that firms are price takers on input markets and that output is exogenously determined seems correct. This appears to be appropriate for a regulated industry where firms are relatively small players on input markets and are required to satisfy market demand at a price set by the regulator.

IV.2 Non-parametric Approach (DEA)

Every DEA model try to determine first which productive units form the simple rest on the envelope surface –the efficient frontier- Then, DEA yields an exhaustive methodology to relative efficiency analysis, evaluating each firm and measuring its performance against the frontier. Productive units yielding on the envelope surface are considered “efficient” in DEA jargon, while the remaining are labeled as “inefficient”, and the analysis provides a measure of the relative (in) efficiency.

The selection of a particular DEA model implies a decision on the shape of the efficient frontier and on the concept of distance (inefficiency) to use. The first decision has to be with an assumption on scale returns. There are basically two alternatives: constant returns to scale (RCE) and variable returns to scale (RVE). The selection of a concept of distance is related to an option for an orientation to the model: to a proportional reduction of the inputs, remaining constant the output level, or to the proportional increase in outputs given the inputs, or none of them.

The theoretical specification of the model RCE oriented to inputs consists in a constrained optimization problem as the following:

$$\begin{aligned}
 [3] \quad & \min_{\theta, \lambda} \theta \\
 & \text{subject to: } y_j \leq \lambda Y \\
 & \lambda X \leq \theta x_j \\
 & \lambda Z = z_j \\
 & \lambda \in R_+^J
 \end{aligned}$$

Where Y is a matrix $N \times r$ of the firm outputs (N denote the number of firms and r the number of outputs); X is a matrix $N \times m$ of inputs (m indexes the inputs considered); Z is a matrix $N \times s$ containing all the information on the S environmental variables from the N firms; y_j , x_j and z_j are the observed vectors of outputs, inputs and environmental variables of the firm under analysis respectively; and, finally, λ is a vector of intensity parameters ($\lambda_1, \lambda_2, \dots, \lambda_N$) which allows the convex combination of the observed inputs and outputs (to construct the envelope surface).

This problem yields as a solution the proportion (θ) in which the observed inputs and the costs of the firms under analysis could be reduced if the firm were efficient.

The efficiency measures just obtained are indexes of productive efficiency, known as Debreu-Farrell measures. Since every measure is the reciprocal of a distance function, the measures cope with some desirable properties. The representation of the technology with distance functions, allow multi-product and multi-input situations, on the contrary as the traditional production functions. So, they avoid the need of adding outputs or inputs before the analysis.

To obtain a RVE model of any orientation, it is just needed to add an additional constraint to the former specification:

$$\sum_j \lambda_j = 1 \quad \forall j = 1, \dots, N.$$

This constraint ensures that an inefficient unit just be compared with productive units of the similar size. Without this constraint, the unit under analysis could be compared with other materially greater or smaller.

Whatever of the precedent problems should be solved N times, once for every firm of the sample. The θ value yielded (one for each productive unit), will be called θ^* (where the ‘*’ denotes an optimal value), is the efficiency measure of the unit under analysis. If the radial contraction of its inputs is possible, $\theta^* < 1$, the productive unit is inefficient and $[(1 - \theta^*) \times 100]$ measures the percentage reduction that could be applied in costs and inputs. For example, if $\theta^* = 0.80$, the productive unit could reduce in a 20% its cost level.

The RVE model is the more desirable option, since it does not constraint the possible scale returns. However there were computed also RCE versions on all the models, since in some cases, considering RVE the smaller and less productive units tend to appear as 100% efficient, simply because they have no comparators.

V. The estimates

V.1. Cost Frontier Estimates

One disadvantage of a cost function estimate is the difficult task of obtaining input price information. With the database for 2005 it could be constructed two measures of price inputs for labor and “capital” (more properly a price index of non-labor inputs). The former was constructed as an average salary in each firm (W) since information on global labor costs and total staff. For “capital”, it was computed the cost of non- labor inputs divided on the total of water clients (R).

Because the database includes firms from different countries, it is worth mention the treatment applied to monetary variables. Local currency information was expressed in American Dollars using the average exchange rate for 2005. To make a correct comparison, all values were converted by purchasing power parities.

It should be recalled that the model structure has two parts: the core of the model, and the environmental variables. To select correctly the latter is a key element to characterize the

context where they operate, to develop fair comparisons. Having these in mind, it was made a detailed selection of both, core and environmental variables.

Between the core variables, there were employed OPEX as dependent, CLIW, W and R. In all cases, monetary variables were expressed in purchasing power parity. And into the environmental, a set of different options were tested; not all of them were finally included in the model.

To determine the model, it was selected the parametric approach (econometrics), since that methodology allows hypothesis testing. The objective was to reach the more robust model or models, accomplishing with the desirable statistical properties.

From the database some variables arisen as good candidates from the theoretical or empirical point of view to explain the phenomena, but the lack of sufficient number of observations or in some cases doubts on the quality of the data suggested us not to be included for now:

- ❑ *Total Costs (there are some problems of consistency)*
- ❑ *Management and commercial costs.*
- ❑ *Outsourced Costs (information on third parties services is not confident at the present date.*
- ❑ *Costs of Power used to operation and maintenance activities.*
- ❑ *Physical units of electricity consumption.*
- ❑ *Differences on regulatory framework, which could isolate qualitative elements acting on costs.*

V.2. Parametric Approach (Econometrics)

The best estimates explain OPEX as depending on CLIW, W and R. As environmental variables there were included UNDW, UNAW, RESW, BREW and COMP, which reflect respectively, the proportion of underground water on total water, the unaccounted for water as a proportion of total production, the residential on total clients, the breaks on water mains divided the main length and the client complains on the network length.

All models were estimated with data from 2005. The specification is a Cobb-Douglas in logarithms. For the stochastic frontier the method used was maximum likelihood (ML), with two alternative assumptions on the residuals: normal/exponential (MLE) and normal/half normal (MLH). The results are presented in the following Table.

The former models are characterized by a two parts estimation error: one regarding on statistic noise and the other being the inefficiency term, which is intending to be isolated.

The models behave in a satisfactory way. The sign of the coefficients is as expected, according to the theory. The variables included are significant, jointly and individually, in order to explain the variability of the OPEX (with the exception of UNDW in MLEB, and RESW in MLHD model).

Since all the variables are expressed in logarithms, the coefficients are to be understood as elasticities, that is, before a percentage change in the explanatory variable under consideration, the coefficients yield the percentage change in OPEX.

Table 4: Alternative Models of Stochastic Frontiers MLE (Normal/Exponential) y MLH (Normal /Half Normal)

Dependent variable:												
OPEX												
Independent Variables: (all in logs)	MLEA	MLHA	MLEB	MLHB	MLEC	MLHC	MLED	MLHD	MLEE	MLHE	MLEF	MLHF
CLIW	0.957*	0.935*	0.981*	0.901*	0.926*	0.905*	0.957*	0.933*	0.877*	0.859*	0.822*	0.807*
	(0.024)	(0.023)	(0.021)	(0.000)	(0.022)	(0.022)	(0.023)	(0.023)	(0.024)	(0.027)	(0.031)	(0.028)
R	0.688*	0.621*	0.697*	0.574*	0.684*	0.605*	0.702*	0.622*	0.597*	0.513*	0.655*	0.625*
	(0.063)	(0.057)	(0.052)	(0.000)	(0.059)	(0.057)	(0.060)	(0.056)	(0.054)	(0.065)	(0.042)	(0.035)
W	0.183*	0.235*	0.156*	0.203*	0.191*	0.264*	0.161*	0.224*	0.082	0.159*	0.039	0.091**
	(0.066)	(0.063)	(0.051)	(0.000)	(0.059)	(0.056)	(0.063)	(0.062)	(0.052)	(0.059)	(0.042)	(0.043)
UNDW			-0.053	-0.101*								
			(0.086)	0.000								
UNAW					0.247*	0.269*						
					(0.067)	(0.077)						
RESW							-0.564***	-0.575				
							(0.332)	(0.415)				
BREW									0.018**	0.020*		
									(0.006)	(0.007)		
COMP											0.096*	0.121*
											(0.022)	(0.026)
CONSTANT	-6.816*	-6.894*	-6.867*	-6.120*	-6.201*	-6.360*	-6.162*	-6.244*	-4.538*	-4.778*	-4.505*	-4.987*
	(0.348)	(0.373)	(0.287)	0.000	(0.353)	(0.391)	(0.494)	(0.599)	(0.476)	(0.569)	(0.422)	(0.441)
Sigma_v	0.220	0.200	0.155	0.000	0.203	0.220	0.204	0.192	0.124	0.140	0.092	0.066
Sigma_u	0.292	0.509	0.322	0.486	0.263	0.412	0.304	0.516	0.291	0.442	0.271	0.426

Note: *** Significant at 10%, **Significant at 5%, * Significant at 1%.

Source: Own elaboration.

The Cobb-Douglas specification could be questioned, because its restrictive assumptions. In this sense, it will be desirable to estimate a trans-logarithmic cost function, but there are not sufficient data to date. It can be an extension for the current study.

Although from the strictly statistical point of view the models satisfy the desirable properties, for its regulatory use more caution is needed, because of the robustness of the model and of its capability to recall relative efficiency

Some work have to be done on data improving and in the inclusion of more variables which properly allow the identification of regulatory differences between countries, more details on input prices, a better proxy for quality of service, variables which address differences in topography, weather, etcetera. The probable empirical relevance of these dimensions is supported by the literature surveyed.

V.3. Relative efficiency measures

Efficiency measures from this model have a reasonable variability between operators, which is good, but they are not definitive measures on efficiency or relative inefficiency

with regulatory application. More observations and more detailed modeling could improve in an important way the latter.

There are four conditions which in principle are desirable in benchmarking studies: number of observations, comparability, a common regulator (or regulation) and information. The database in use does not accomplish to date with all those properties.

The current amount of observation is a good number, but the blanks in the database in precedent years prevent us to estimate a panel; instead the study has been made based on a cross section of the year 2005.

With respect to the issue of comparability, the firms are located in different countries, cities and regions, each one very different between the others. Population density, technology in use and a set of additional dimensions are to be considered. However, the inclusion of environmental variables has ameliorated the task. The conversion of monetary values to purchasing power parity is another effort in the direction to made fairer comparisons.

Regarding to a common regulator, that condition is not accomplished. However, it could be sufficient with the inclusion of some measures, which can proxy that different regulatory environment, as quality standards, the character of public o private firms, etcetera. To incorporate more precisely the differences in regulation, it will be useful a comparative study of the regulations which could allow extracting qualitative details. The incorporation of these details exceeds the scope of this paper, since the data have not been produced yet.

The fourth condition will be solved with the continue improvement of the ADERASA database and the addition of new observations.

In the Table 5 the descriptive statistics of the efficiency measures yielded from the former models are shown. It was made a selection of models according to the individual variables and also the selection was made resting on tests, which could yield variability between the inefficiency and statistic noise (in the case of stochastic frontiers). Some of those models have the same specification, but different assumptions on the statistic distribution of the inefficiency. The remaining models are different alternatives including environmental variables.

Estimates with the year 2005 data, show differences between minimum and medium values, and the standard deviations have variability, which are acceptable in the normal practice of these kinds of measures.

Table 5: Descriptive Statistics of the Efficiency Measures

Model	N	Average	Std. Dev.	Min	Max
MLEA	72	0.760	0.149	0.126	0.945
MLEAD	71	0.756	0.157	0.112	0.949
MLHA	72	0.683	0.153	0.166	0.934
MLHC	72	0.728	0.124	0.260	0.929
MLHE	47	0.725	0.154	0.316	0.945
MLHf	53	0.736	0.165	0.287	0.967

However, in the case of stochastic frontiers with different distributions, the levels of efficiency are very similar. But, in this case the half-normal model is less efficient than the exponential model.

V.4. Non-Parametric Approach (DEA)

As it can be seen in the formulation of the problem, it was chosen to model the environmental variables (z_j) as neutral, not discretionary variables (under which the firm has no control). In this option, every firm is only compared against a hypothetical firm, which operates in the same environmental as the firm under evaluation. The main advantage of this option is that it does not imply an a priori judgment on the sense of the influence of every environmental variable over the efficiency.

We choose an approach of resource conservation (input oriented), and it was resorted to a group of models presented in the Table 6. As it is develop a cost efficiency model, the only input that was used is the OPEX. As output it was considered in all cases the water clients which, as mentioned in the econometric estimates, has a high correlation with other possible output measures, like the volume of water produced or the coverage.

Table 6: DEA. Model specifications and relative efficiency results

	RCEA	RCEB	RCEC	RCED	RCEE	RVEA	RVEB	RVEC	RVED	RVED
Outputs	CLIW	CLIW	CLIW	CLIW	CLIW	CLIW	CLIW	CLIW	CLIW	CLIW
Inputs	OPEX	OPEX	OPEX	OPEX	OPEX	OPEX	OPEX	OPEX	OPEX	OPEX
Environment	None	UNDW	UNAW	COVN	UNDW UNAW COVN	None	UNDW	UNAW	COVN	UNDW UNAW COVN
Orientation	Inputs	Inputs	Inputs	Inputs	Inputs	Inputs	Inputs	Inputs	Inputs	Inputs
Returns to scale	Constants	Constants	Constants	Constants	Constants	Variables	Variables	Variables	Variables	Variables
Average	0.263	0.367	0.316	0.339	0.620	0.323	0.518	0.441	0.442	0.751
Std. Dev.	0.149	0.211	0.206	0.225	0.297	0.211	0.281	0.273	0.269	0.278
Min	0.056	0.093	0.064	0.077	0.122	0.093	0.103	0.117	0.110	0.216
Max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The Table 6 also shows the different alternatives chosen. The models with constant returns to scale RCEB, RCEC and RCED add each one a different environmental variable. The RCEE model includes the set of the three variables already included on an individually basis. Finally, it were computed the same models with variable returns to scale (RVE).

As it can be seen, the average of the efficiency measures is greater in the RVE model, which is logical due to the returns assumption. This model tends to identify a greater number of firms as relatively more efficient. The variability of results, in turn, is reasonable.

V.5. Consistency analysis

In a tariff review, the initial regulatory task implies to assess if the productivity gains used to set the new price cap is specific to the firm and if it is based on past profits. If it is so, the firm will not have powerful incentives to cost reductions, since it yields a lower tariff. An alternative for the regulator is to measure the efficiency gains in a way such they do not rest

under the control of the firm, as it is done in yardstick competition, in which the prices could be set with reference to the aggregate performance of the industry.

If a firm has a cost efficiency measure of 0.8, for instance, it means that it could produce the same output level with the 80% of its current costs: there are firms doing that, (those with cost efficiency measure of 1 or 100%). This implies that the price cap should be based on the 80% of the current costs, not on the 100%. Following this way, only the 100% efficient firms could recover its capital opportunity costs, while the others will have a lower return.

In this context, the efficiency assessment at the level of the firms depends both on the methodology employed and on the selection of the explanatory variables. The regulators applying benchmarking could have problems if the different possible analysis yields contradictory results. One solution is the consistency analysis. This poses certain basic conditions, which should be accomplished in order to assess the utility of the results for the regulatory authorities. The advantage of a consistency analysis is that it does not turn the methodology choice mandatory. The regulator could avoid the choice, and instead of that, could use many techniques and cross check the different results.

The consistency conditions, as Bauer et al (1998) suggested, demand that the different methodologies:

- (i) Yield similar distributions of the efficiency measures;
- (ii) Generate alike unit rankings;
- (iii) Identify the same units as the “best” and the “worst”;
- (iv) Produce efficiency measures stable in time;
- (v) Be reasonably consistent with other performance measures (as those of partial productivity);
- (vi) Are consistent with the conditions which the industry as a whole faces.

To implement a mechanism as that suggested, demands at least the accomplishment of the first consistency condition (similar distributions of the efficiency measures). If that condition is not reached, the mechanism should not be applied, because the measures seem in some sense to be subjective and non confident.

In general, it could be said that the first consistency condition is partially achieved and that the different between averages are due to model specification mainly. It can be seen that the efficiency measures obtained from econometric models do not differ in a significant way (Table 5). For DEA, the efficiency average increases when more environmental variables are added (Table 6). The efficiency averages obtained with DEA are lower than those obtained with econometric estimations.

If the efficiency levels are not consistent between the different models, yet it is possible that the same generate similar rankings of the firms according to its efficiency measures, accomplishing the second consistency condition, which it turn could help to discriminate the X-factor between firms. The Tables 10 and 11 show the Spearman correlation coefficient between rankings established by pairs of models for both specifications. All the correlations are positive and significantly different from zero at the 5% level. The same

indicate the similarity of the intra-methodology results. Correlations are higher between the non-parametric models. This result is also found in the applied literature, where the intra-methodology consistency seems to be greater than the inter-methodology consistency. In general, the evidence is not conclusive on the accomplishment of this condition, although the results point to some robustness of the conclusions.

Table 10: Spearman's Correlation Coefficients. Econometrics: selected models

Correlation coefficients						
	MLEA	MLED	MLHA	MLHC	MLHE	MLHF
MLEA	1.000					
MLED	0.879	1.000				
MLEHA	0.905	0.879	1.000			
MLHC	0.908	0.854	0.911	1.000		
MLHE	0.536	0.514	0.532	0.513	1.000	
MLHF	0.744	0.721	0.747	0.697	0.489	1.000

Table 11: Spearman's Correlation Coefficients. DEA. Selected models.

Correlation coefficients										
	RCEA	RCEB	RCEC	RCED	RCEE	RVEA	RVEB	RVEC	RVED	RVEE
RCEA	1.000									
RCEB	0.879	1.000								
RCEC	0.927	0.888	1.000							
RCED	0.948	0.839	0.957	1.000						
RCEE	0.445	0.646	0.552	0.509	1.000					
RVEA	0.935	0.880	0.991	0.967	0.554	1.000				
RVEB	0.701	0.782	0.752	0.721	0.668	0.778	1.000			
RVEC	0.720	0.729	0.783	0.762	0.636	0.800	0.651	1.000		
RVED	0.707	0.744	0.814	0.799	0.676	0.811	0.647	0.738	1.000	
RVEE	0.300	0.444	0.383	0.334	0.706	0.385	0.587	0.599	0.460	1.000

To summarize, the first consistency condition is verified. Econometric models present efficiency measures greater than those obtained with DEA. This is a fact present in the literature. The inter-methodology results are better than the intra-methodology ones. The comparison between DEA-DEA models identifies consistent best (and worst) practices. Results are less satisfactory when DEA and econometrics are compared.

VI. Conclusions

The aim of this study is to estimate both stochastic and mathematical programming cost functions and frontier studies with the database developed by ADERASA. Econometric and DEA techniques were employed to the calculations. The study fills a gap on the understanding of Latin American water and sanitation sector relative efficiency.

The econometric estimates were developed firstly, as a guide for the construction of DEA models later. There were studied different proxies of outputs, inputs, input prices and environmental variables. The monetary variables were converted to purchasing power parity values. Salaries and an index of the non-labor costs were the input prices included. In both cases, the values result from processing global amounts from the ADERASA survey, in the line of the literature practice.

On the efficiency concept to measure, there were developed estimates of Cost functions, which represent the total cost of production depending on output level input prices and environmental variables. They allow estimating productive or total efficiency.

The methodologies used were econometrics and DEA. Regarding the econometric approach, it could be concluded that the estimates are satisfactory. The signs of the coefficients are as expected, according to the theory. The variables included are significant jointly and individually to explain the variability of the OPEX at the usual confidence levels. There were found also significant environmental variables. With the 2005 data, the estimates resulted robust in many different specifications

Once estimated the econometric models, the following step was applying DEA method to the same set of variables (output, input prices and environmental). There were considered two possibilities: constant returns to scale (RCE) and variable returns to scale (RVE).

A common problem faced by regulators applying benchmarking is the great amount of disposable methodologies as the various alternatives of modeling. The problem is more acute if the different methodologies offer contradictory results. To cope with these problems, a consistency analysis on the results is needed.

For the estimates presented in the paper, an internal consistency analysis was performed, concluding that at a general level the intra-methodology comparison yields similar results.

References

- Berg, Sanford and Chen Lin (2005). "Consistency in Performance Rankings: The Peru Water Sector." May 6.
- Bauer, P., A. Berger, G. Ferrier, y D. Humphrey, 1998. "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods". *Journal of Economics and Business* 50, pp. 85-114.
- Bosworth, D., P. Stoneman and E. Thanassoulis (1996). "The measurement of comparative total efficiency in the sewerage and water industry: An exploratory study," Report commissioned by the Office of Water Services, October.
- Bottasso, A. and M. Conti (2003). "Cost Inefficiency in the England and Welsh Water Industry: A Heteroskedastic Stochastic Cost Frontier Approach", *Economics Discussion Papers*, University of Essex, Department of Economics.
- Bosworth, Derek, Paul Stoneman and Emmanuel Thanassoulis (1996). "The Measurement of Comparative Total Efficiency in the Sewerage and Water Industry: An Exploratory Study". A Report to and Commissioned by the Office of Water Services (OFWAT).

- Clarke, George R.G., Katrina Kosec, and Scott Wallsten (2004). "Has Private Participation in Water and Sewerage Improved Coverage? Empirical Evidence from Latin America," Working Paper 0402 (January), AEI-Brookings Joint Center for Regulatory Studies.
- Crampes, C, N. Niette and A. Estache (1997). "What Could Regulators Learn from Yardstick Competition? Lessons for Brazil's Water and Sanitation Sector. Mimeo. The World Bank.
- Estache, Antonio and Martín Rossi (1999). Estimación de una frontera de costos estocástica para empresas del sector agua en Asia. Serie Textos de Discusión N° 4. CEER/UADE. Buenos Aires, abril.
- Estache, Antonio and M. Rossi (2002). "Comparing the Performance of Public and Private Water Companies in Asia and Pacific Region. What a Stochastic Costs Frontier Shows". Policy Research Working Paper. Washington, The World Bank.
- Estache, Antonio and Eugene Kouassi (2002). Sector Organization, Governance, and the Inefficiency of African Water Utilities. Policy Research Working Paper 2890. Washington, The World Bank.
- Fraquelli, Giovanni and Valentina Moiso (2005). Cost Efficiency and Economies of Scale in the Italian Water Industry. XVII Conferenza Società Italiana di Economia Pubblica. Pavia, Università, 15-16 settembre 2005.
- Gómez-Lobo, A., 2004. "Determinación de la eficiencia operativa en la regulación de monopolios naturales: el uso de información de consultores versus competencia por comparaciones," Departamento de Economía, Universidad de Chile, Octubre.
- Lin, Chen (2005). "Service Quality and Prospects for Benchmarking: Evidence from the Peru Water Sector", Public Utility Research Center, University of Florida, P. O. Box 117142, Gainesville, FL 326117142. Revised April 25.
- Margaretic, P., D. Petrecollo y C. Romero, 2004. "Evaluación de la eficiencia de la operación de las distribuidoras de electricidad en la Argentina. Un análisis ex-post de la década de los 90," Fundación Fines, Buenos Aires, Agosto
- Mobbs, Peter and Edward Glennie (2004). "Econometric modelling with ADERASA data & indicator values", WRc (Draft).
- Moreira, Ajax R. B. and Thais C. R. Fonseca (2005). "Comparando Medidas de Productividad: DEA, Frontera de Producción Estocástica," Texto para discusión 1069, IPEA, Feb.
- Ofwat (2002). Water and Sewerage Services Unit Costs and Relative Efficiency, Birmingham, Office of Water Services.
- Price, Jonathan (1993). "Comparing the Cost of Water Delivered. Initial Research into the Impact of Operating Conditions on Company Costs". Costs and Performance Division. OFWAT. Birmingham (March).
- Romero, Carlos (2005). Benchmarking de empresas de agua y saneamiento de Latinoamérica sobre la base de datos de ADERASA. Informe Final. Paper prepared for ADERASA.
- Sabbioni, Guillermo (2005). Econometric Measures of the Relative Efficiency of Water and Sewerage Utilities in Brazil. Public Utility Research Center, University of Florida, May 10.
- Saal, David and David Parker (2005). Assessing the Performance of Water Operations in the English and Welsh Water Industry: A Panel Input Distance Function Approach. Aston Business School Research Paper RP0502. Aston University. Birmingham, January.

- Saal, D. and Scott Reid (2004). "Estimating Opex Productivity Growth in English and Welsh Water and Sewerage Companies: 1993-2003," Aston Business School Working Paper RP0434.
- Saal, David, David Parker and Tom Weyman-Jones (2004). Determining the Contribution of Technical Efficiency and Scale Change to Productivity Growth in the Privatized English and Welsh Water and Sewerage Industry: 1985-2000. Aston Business School Research Paper RP0433. Aston University. Birmingham, December.
- Saal, D. and D. Parker, 2001. "Productivity and Price Performance in the Privatized Water and Sewerage Companies of England and Wales," *Journal of Regulatory Economics* 20:1, pp. 61-90
- Stewart, Mark (1993). "Modelling Water Costs 1992-93. Further Research into the Impact of Operating Conditions on Company Costs". OFWAT. Birmingham.
- Stewart, Mark (1994). "Modelling Sewerage Costs 1992-93. Research Into the Impact of Operating Conditions on the Costs of the Sewerage Network". OFWAT. Birmingham. (1993)
- Tupper, Henrique Cesar, and Marcelo Resende (2004). "Efficiency and Regulatory Issues in the Brazilian Water and Sewage Sector: An Empirical Study," *Utilities Policy* 12, pp. 29-40.

CEER Working Paper Series

To order any of these papers, or all of these, see instructions at the end of the list. A complete list of CEER Working Papers is displayed here and in our web site

- WPS 1. *Laffont, Jean Jacques: Translating Principles Into Practice in Regulation Theory (March 1999)*
- WPS 2. *Stiglitz, Joseph: Promoting Competition in Telecommunications (March 1999)*
- WPS 3. *Chisari, Omar, Antonio Estache, y Carlos Romero: Winners and Losers from Utility Privatization in Argentina: Lessons from a General Equilibrium Model (March 1999)*
- WPS 4. *Rodríguez Pardina, Martín y Martín Rossi: Efficiency Measures and Regulation: An Illustration of the Gas Distribution Sector in Argentina (April 1999)*
- WPS 5. *Rodriguez Pardina, Martín Rossi and Christian Ruzzier: Consistency Conditions: Efficiency Measures for the Electricity Distribution Sector in South America (June 1999)*
- WPS 6. *Gordon Mackerron: Current Developments and Problems of Electricity Regulation in the European Union and the United Kingdom (November 1999)*
- WPS 7. *Martín Rossi: Technical Change and Efficiency Measures: The Post-Privatisation in the Gas Distribution Sector in Argentina (March 2000)*
- WPS 8. *Omar Chisari, Martín Rodriguez Pardina and Martín Rossi: The Cost of Capital in Regulated Firms: The Argentine Experience (May 2000)*
- WPS 9. *Omar Chisari, Pedro Dal-Bó and Carlos Romero: High Tension Electricity Network Expansions in Argentina: Decision Mechanisms and Willingness-to-Pay Revelation (May 2000).*
- WPS 10. *Daniel A. Benitez, Antonio Estache, D. Mark Kennet, And Christian A. Ruzzier. Potential Role of Economic Cost Models in the Regulation of Telecommunications in Developing Countries (August 2000).*
- WPS 11. *Martín Rodríguez Pardina and Martín Rossi. Technical Change and Catching-up: The Electricity Distribution Sector in South America (November 2000).*
- WPS 12. *Martín Rossi and Iván Canay. Measuring Inefficiency in Public Utilities: Does the Distribution Matter? (April 2001).*
- WPS 13. *Quesada, Lucía. Network Competition and Network Regulation (July 2001).*
- WPS 14. *Rossi, Martín and Christian Ruzzier: Reducing the Asymmetry of Information Through the Comparison of the Relative Efficiency of Several Regional Monopolies (July 2001).*
- WPS 15. *WPS 15. Ferro, Gustavo: Political Risk and Regulatory Risk: Issues in Emerging Markets Infrastructure Concessions (August, 2001).*
- WPS 16. *Margaretic, Paula, María Fernanda Martínez and Diego Petrecolla: The effectiveness of Antitrust Enforcement in Argentina, Chile and Perú during the 90' (March, 2005).*
- WPS 17. *WPS 17. Chisari, Omar, Antonio Estache, Germán Lambardi and Carlos Romero: Trade balance effects of infrastructure services liberalization and of their regulation (March, 2005).*
- WPS 18. *Chisari, Omar and Lucía Quesada: Trade balance constraints and optimal regulation (March 2005)*

- WPS 19. *Ferro, Gustavo: Options for Radical Reforms to Pension Systems: Chilean and Swedish Models Compared (April 2005)*
- WPS 20. *Del Aguila, Virginia: Establishing global competition standards: Achievable Mission or Utopia? (April 2005)*
- WPS 21. *Ferro, Gustavo and Carlos Romero: A cost frontier for Pension Funds in Argentina. (April 2006)*
- WPS 22. *Ferro, Gustavo and Carlos Romero: Efficiency in water and sanitation sector. A survey on empirical literature (May 2007)*
- WPS 23. *Chisari, Omar: A note on Access Pricing, Role Exchangeability and Incentives to Invest (May 2007)*
- WPS 24. *Ferro, Gustavo, Javier Maquieyra and Carlos Romero: On the regulatory choice of returns to scale in DEA MODELS: An application to the argentine electricity distribution sector (June 2007)*
- WPS 25. *Chisari, Omar and Martín Cicowiez: On the Marginal Cost of Public Funds for Argentina: CGE Evaluation and Sensitivity to Regulatory Regimes (May 2008)*
- WPS 26. *Ferro, Gustavo and Carlos Romero: Are There Cost Differences in the Argentinean Pension Fund Industry? An Efficiency Frontier Analysis. (May 2008)*
- WPS 27. *Romero Carlos and GustavoFerro: A Benchmarking Exercise on Latin American Water Utilities (August 2008)*